

意味・文脈・発話権情報を用いた音声対話用文解析手法

藤江 真也 神作 洋一 宮永 悠平 小林 哲則

早稲田大学理工学部

{fujie,ykansaku,miyanaga,koba}@tk.elec.waseda.ac.jp

1 はじめに

人間と自然な対話を実現する音声対話への応用を目標として、以下の二つの点を考慮した言語処理手法を検討する。

まず第一に、入力が多少の誤りを含んでいたとしても、対話の流れを崩すことなく動作することである。音声対話システムの入力である音声による発話は、口語調で崩れた文であることや、音声認識の誤りから、誤りを含むことが多い。しかし、これらの誤りを含んだ文も、構文的には間違っているということが多い。この様な入力をそれなりに解析する手法が必要である。

第二に、入力を逐次的に処理することである。人間同士の音声対話を考えると、一方の発話が終了する前にもう一方が話し始めるという現象が見られる。このことから、人間は逐次的に相手の発話を理解していることが伺える。また、自然な対話をするためには、人間が発話を終了してからできるだけ早く応答を返すことが望ましく、そのためにも逐次的な処理が必要であるとも言える。

本研究では、構文解析時に意味・文脈・発話権情報を考慮することによって、上で挙げた二つの点を考慮した音声対話用文解析手法を提案する。本手法は、プロ野球を話題とした雑談に参加できる対話システムを目指して開発されているROBITA(Real world Oriented BI-modal Talking Agent)[1]上の音声対話システムの一部として実装されている。対話風景を図1に示す。

構文解析と意味解析を同時に行なうシステムは従来から存在する[2][3]が、これらは文脈解析を同時に行なわない。また、文脈を用いて逐次的に処理するシステムも存在する[4]が、本研究で対象とする雑談のような対話においては、文脈を各話者毎に用意する必要があると考えられる。

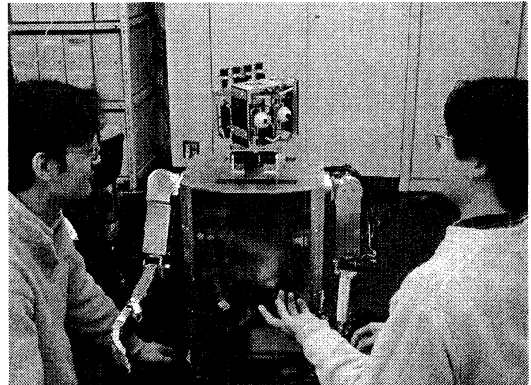


図 1: 対話風景

2 音声対話用文解析手法

本研究で提案する手法は、従来の言語処理で別々に処理されていた構文解析、意味解析、文脈解析を構文解析で統一的行なう。

構文解析の主な処理は、複数の句を一つに組み上げることである。組み上げる基準には構文規則を利用するが、正しい文を対象とした構文規則では崩れた文には対処できない。そこで、構文規則を崩れた文を許すように拡張することが考えられるが、これにより曖昧性が増大し計算量が爆発すると考えられる。そこで、本手法では句を一つに組み上げる際に意味的な制約も考慮し、助詞などの機能語の欠落や誤りに対処すると同時に、句の意味を計算する。

また、音声対話における発話は、主語や目的語、述語が多く省略される。人間は、これらが対話が行なわれている状況やそれまでの対話の流れによって補完できる場合はいちいち発話しないからである。このような省略を含む入力に対して、構文解析の前処理的に予め文脈から意味を補完すること

により、構文解析の複雑さを抑えられる。

さらに、音声対話システムの流れを円滑にするために、発話権を考慮する。

以下では、各情報の考慮に関して順に説明する。

2.1 意味情報の利用

本研究で用いる意味モデルは、主に名詞の意味を表すカテゴリと、主に動詞の意味を表す関係に分かれる。これらはタスクに合わせて定義した。カテゴリは属性、関係は役割をそれぞれ持ち、それらに当てはまるカテゴリを定義しておく。

意味を考慮した構文解析の例を図2に示す。図中、四角で囲まれたものが意味モデルによって定義された、語または句の意味である。図のように、「ホームランを」と「打つ」という句を組み上げる際に、「ホームラン」に「打つ(x,y,z)」の役割yが与えられる。役割が与えられる際に用いる基準として、助詞「を」を伴っていることを用いるだけでなく、役割がカテゴリ「ヒット」に制限されていることも用いる(カテゴリ「ホームラン」はカテゴリ「ヒット」の子クラスなので、この制約を満たす)。これにより、助詞「を」が欠落、もしくは誤っていたとしても正しく処理することが可能である。

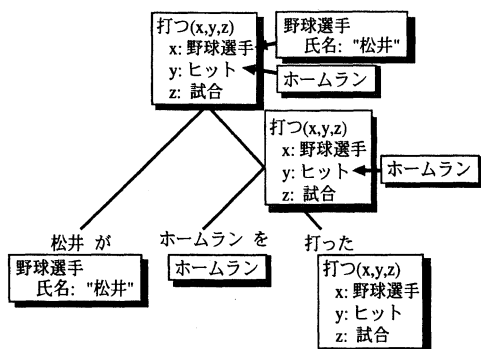


図2: 意味を考慮した構文解析の例

2.2 文脈情報の利用

本研究では、ユーザの知識を文脈として利用する。ここで言うユーザの知識は、システムがユーザの知識に対して持っている信念に過ぎない。従って、ユーザの持つある知識を文脈として用いる場

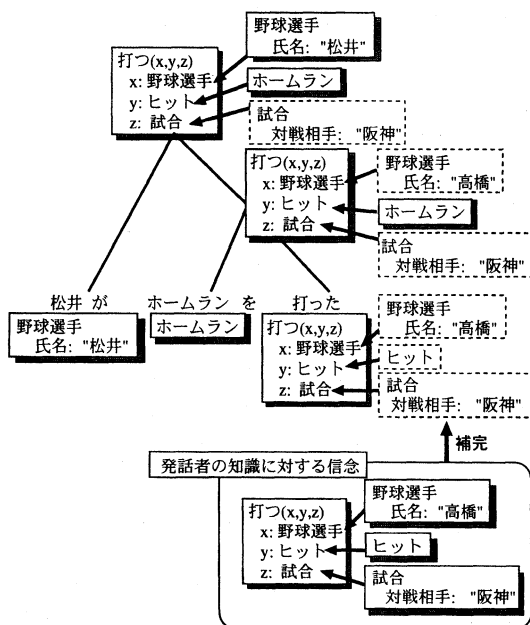


図3: 文脈を考慮した構文解析の例

合に、どの程度信用できるかを考慮しなければならない。

ユーザがある知識を持っているとシステムが判断する理由を、次の三つのパターンに分類した。

1. 直接、知識を与える発話を受けた。
2. 間接的に知識を与える発話を聞いた。
3. 事前に知っていた。

それぞれに一定の信頼度を与える。

文脈を考慮した構文解析の例を図3に示す。この例では、発話者であるユーザが「高橋が阪神戦でヒットを打った」という事実を知っていると、システムは信じている。この場合の「松井がホームランを打つ」という入力に対する処理を示している。

「打つ」という句の持つ意味に対して、文脈である発話者の知識から、役割を与えられる個体の補完を試みる。例ではそれぞれ「高橋」、「ヒット」、「阪神戦」が補完されている。補完された個体に関しては、これが補完されたものであるという印を付けておく(図では点線の四角でそれを表している)。その後の処理で、今回の入力から役割を与えることのできる個体があれば、そちらを優先する。

2.3 発話権の利用

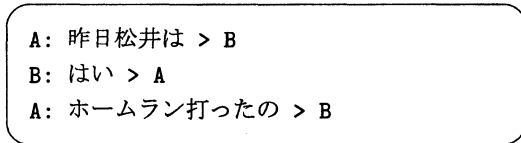


図 4: 発話が複数に分かれている例

音声対話では、まとまった意味を伝えるために発話を何度かに分けることがよく行われる。例えば図 4 のような対話において、ユーザ A の発話の処理を考えると、先の発話を文脈として後の発話を処理するよりも、両者をまとめて処理した方がよいことが伺える。

本手法では、まとめて処理すべき単位を発話権の移動を元に決定する。ユーザの発話による発話権の移動情報は、文頭・文末表現から決定する。例えば、「は」で終わる表現は発話権の維持を表し、「たの」で終わる表現は発話権の譲渡を表す。また、「はい」などの相槌は相手の発話権の維持の承認を表す。

図 4 の対話では発話権は A が保持しているが、最後の発話が発話権の譲渡を表すので、ここで初めて発話権が移動する。この時点で A の発話に関する構文解析結果をまとめることにより、まとまった意味を得る。

3 システム構成

図 5 にシステムの概観を示す。システムの入力は音声認識結果である単語列と、発話者・受話者認識結果の発話者・受話者情報である。発話者・受話者認識は顔方向認識などの結果を用いる。一方、出力は応答を生成する為の意味である。

発話権管理部は、単語列と発話者・受話者情報から発話権を計算し、構文解析部に出力する。行動決定部は、構文解析結果である意味構造と発話者・受話者情報からシステムの行動を決定する。発話がシステムに向けられたもので、また発話意図が情報要求だった場合、ユーザに応答するため、意味構造を応答生成に出力する。また、意味構造から

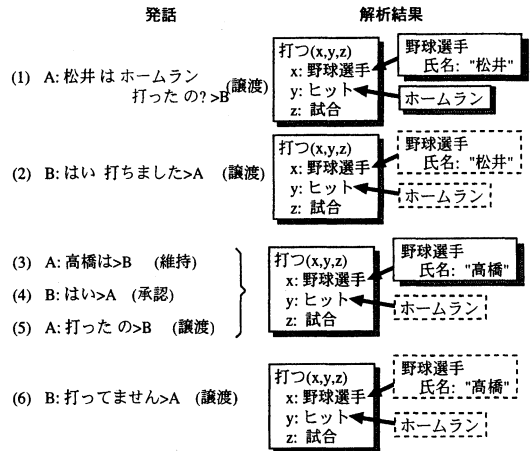


図 6: 動作例

発話者が知っていると判断できる意味構造をユーザ知識管理部に出力し、ユーザの知識を更新する。

システムの持つ知識は、WWW 上にあるプロ野球に関する情報を DB 化したものである。基本はオフライン(対話が始まる以前)に WWW にアクセスし、情報を DB 化するが、DB に無い情報が話題に上がった時などはその時点で WWW にアクセスする機能を持つ。

4 動作例

システムの動作例(ユーザ同士の対話の解析例)を図 6 に示す。入力は、(1)~(6)の横に

発話者：単語列 > 受話者

という形式で書かれている。入力の横に括弧付きで書かれているのは発話権の移動情報である。その横に書かれているのが、発話の解析結果である。

(3) の発話は、発話権の移動情報が「維持」であるので、発話権が移動しない。したがって、構文解析結果は出力されずに保留される。続く(4)で「承認」が行われた後、(5)の発話終了時に発話権が移動するので、ここで(3)の続きとして処理を続け、結果を出力する。

(2)~(6)の発話の解析結果は、それ以前の発話によって更新されたユーザの知識を元に補完された

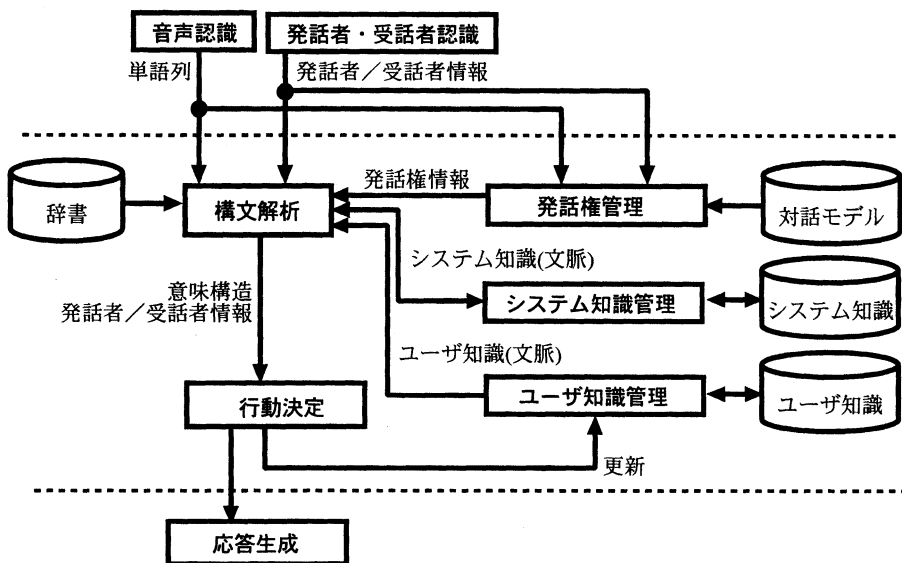


図 5: システム概観

部分を含んでいるのが分かる。

5 まとめ

音声対話用文解析手法として、意味・文脈・発話権情報を用いて構文解析を行なう手法を提案した。本手法は、従来の手法に比べ、助詞などの機能語の間違いや欠落に強く、多くの省略を含んだ発話に対する意味の補完によって、構文解析を行なう前に意味的な曖昧性を減少させることができる。また、発話権を考慮することで、適切なタイミングで出力を行うことができる。

残された課題としては、構文情報、意味情報、文脈情報の考慮に関して、信頼度を導入することが挙げられる。これにより、各情報のうち、どの情報をより信頼することができるか、処理の結果得られた複数の結果のうち、どの結果がよりもらしいか、などの情報を得ることができる。また、韻律やタイミングを考慮することにより、より効率的な構文解析を行うことができると考えられる。また、これらの手法の評価法の確立も重要な問題と考えられる。

参考文献

- [1] Y. Matsusaka, T. Kobayashi, "Multi-person conversation via multi-modal interface - A robot who communicate with multi-user", *Proc. of Eurospeech*, Vol.4, pp.1723-1726, 1999.
- [2] 竹林, "音声自由対話システム TOSBURG II - ユーザ中心のマルチモーダルインターフェースの実現に向けて -", *信学論*, Vol.J77-D-II, No.8, pp.1417-1428, 1994.
- [3] H. Asoh, T. Matsui, J. Fry, F. Asano, and S. Hayamizu, "A Spoken Dialog System for a Mobile Office Robot", *Proc. of Eurospeech'99*, pp.1139-1142, 1999.
- [4] 中野, 宮崎, 堂坂, 平沢, 安田, 川端, "多重文脈を用いた逐次的な発話理解", *情処技報*, 98-SLP-22, pp.21-26.
- [5] 菊池, 工藤, 小林, 白井, "音声対話インターフェースにおける発話権管理による割込みへの対処", *信学論* Vol. J77-D-II, No.8, pp.1502-1511, 1994.